



# BMJ Open Cohort profile: the MCC-Spain follow-up on colorectal, breast and prostate cancers: study design and initial results

Jessica Alonso-Molero <sup>1</sup>, Antonio J Molina,<sup>2</sup> Jose Juan Jiménez-Moleón,<sup>3,4</sup> Beatriz Pérez-Gómez <sup>3,5,6</sup>, Vicente Martín,<sup>2,3</sup> Victor Moreno,<sup>3,7,8</sup> Pilar Amiano,<sup>3,9</sup> Eva Ardanaz,<sup>3,10,11</sup> Silvia de Sanjose,<sup>3,7</sup> Inmaculada Salcedo,<sup>3,4</sup> Guillermo Fernandez-Tardon,<sup>3,12</sup> Juan Alguacil,<sup>3,13</sup> Dolores Salas,<sup>3,14</sup> Rafael Marcos-Gragera,<sup>3,15</sup> Maria Dolores Chirlaque,<sup>3,16</sup> Nuria Aragonés,<sup>3,17</sup> Gemma Castaño-Vinyals,<sup>3,18,19,20</sup> Marina Pollán,<sup>3,5,6</sup> Manolis Kogevinas,<sup>3,18,19,20</sup> Javier Llorca<sup>1,3</sup>

**To cite:** Alonso-Molero J, Molina AJ, Jiménez-Moleón JJ, *et al.* Cohort profile: the MCC-Spain follow-up on colorectal, breast and prostate cancers: study design and initial results. *BMJ Open* 2019;**9**:e031904. doi:10.1136/bmjopen-2019-031904

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-031904>).

JA-M, AJM and JJJ-M contributed equally.

Received 24 May 2019  
Revised 10 September 2019  
Accepted 24 October 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Jessica Alonso-Molero; [alonsoamoleroj@gmail.com](mailto:alonsoamoleroj@gmail.com)

## ABSTRACT

**Purpose** Since 2016, the multicase-control study in Spain (MCC-Spain) has focused towards the identification of factors associated with cancer prognosis. Inception cohorts of patients with colorectal, breast and prostate cancers were assembled using the incident cases originally recruited.

**Participants** 2140 new cases of colorectal cancer, 1732 of breast cancer and 1112 of prostate cancer were initially recruited in 12 Spanish provinces; all cancers were incident and pathologically confirmed. Follow-up was obtained for 2097 (98%), 1685 (97%) and 1055 (94.9%) patients, respectively.

**Findings to date** Information gathered at recruitment included sociodemographic factors, medical history, lifestyle and environmental exposures. Biological samples were obtained, and 80% of patients were genotyped using a commercial exome array. The follow-up was performed by (1) reviewing medical records; (2) interviewing the patients by phone on quality of life; and (3) verifying vital status and cause of death in the Spanish National Death Index. Ninety-seven per cent of recruited patients were successfully followed up in 2017 or 2018; patient-years of follow-up were 30 914. Most colorectal cancers (52%) were at clinical stage II or lower at recruitment; 819 patients died in the follow-up and the 5-year survival was better for women (74.4%) than men (70.0%). 71% of breast cancers were diagnosed at stages I or II; 206 women with breast cancer died in the follow-up and the 5-year survival was 90.7%. 49% of prostate cancers were diagnosed at stage II and 32% at stage III; 119 patients with prostate cancer died in the follow-up and the 5-year survival was 93.7%.

**Future plans** MCC-Spain has built three prospective cohorts on highly frequent cancers across Spain, allowing to investigate socioeconomic, clinical, lifestyle, environmental and genetic variables as putative prognosis factors determining survival of patients of the three cancers and the inter-relationship of these factors.

## Strengths and limitations of this study

- 4837 incident cases of cancer (2097 colorectal; 1685 breast; 1055 prostate) have been prospectively followed up, accounting for more than 30 000 patient-years and with only 153 patients (3%) lost to follow-up.
- The cohort covers a wide spectrum of the Spanish population including 23 hospitals across Spain.
- A major strength of this study is the amount of information gathered at diagnosis, including sociodemographic, lifestyle, nutrition, familial and personal medical history, reproductive history, use of drugs, sleep, genotyping, clinical and pathological characteristics of the tumour, first-line treatment, side effects, health-related quality of life, and current vital status.
- Biological samples obtained at recruitment (tumour specimen, blood or saliva, toenail, hair and urine) will allow further investigations on metabolomics, epigenetics and exposure to chemicals such as metals.
- The multicentre characteristic of the study allows the evaluation of a wide geographical basis and increases the representativeness of the recruited sample, but it also may introduce heterogeneity in the information gathered and in treatment.

## INTRODUCTION

Tumour size, node infiltration, metastasis, histology, clinical stage and cancer subtype continue to be the main prognosis factors in patients with cancer in spite of the evolving first-line treatment.<sup>1–5</sup> Little effort, however, has been paid to examine the impact on survival of patient factors, such as lifestyle, genetics or environmental, together with tumour features and treatment.

Large prospective cohort studies on cancer focus on identifying risk factors,<sup>6</sup> while



clinical cohorts on cancer survival usually aim to analyse survival relationships with tumour properties, first-line treatment or patient characteristics. For instance, Lagendijk *et al*<sup>7</sup> analysed data on 129 692 women with breast cancer from the Netherlands Cancer Registry to compare breast conserving therapy and mastectomy in subgroups according to age at diagnosis, stage, systemic therapy, comorbidity, oestrogen/progesterone receptors and Her2 status. Cardwell *et al*<sup>8</sup> linked the National Cancer Data Repository to the UK Clinical Practice Research Datalink and mortality data from the Office for National Statistics to investigate if statin use after colorectal cancer diagnosis was associated with better prognosis. Pettersson *et al*<sup>9</sup> studied survival after prostate cancer diagnosis in 121 392 Swedish men from the Prostate Cancer data Base Sweden V.3.0, where data were available on age, stage, grade, prostate-specific antigen (PSA) level, model of detection, comorbidity, educational level and primary treatment.<sup>9</sup> It is noteworthy that these cohorts were based on cancer registries, where data availability is usually restricted to demographic variables (sometimes including educational level and deprivation), tumour characteristics and few data on comorbidities or healthy habits. A different approach has been the use of the Surveillance, Epidemiology and End Results (SEER) database to retrospectively analyse survivorship with breast cancer,<sup>10</sup> colorectal cancer<sup>11</sup> or prostate cancer,<sup>12</sup> but although the number of participants could be over 100 000, available data are restricted to those recorded for the general purposes of the SEER programme, not specifically for studying survivorship with cancer.

The multicase-control study in Spain (MCC-Spain) includes three prospective cohorts of patients with cancer (colorectal, female breast and prostate) with the aim to investigate long-term survival factors, including cancer characteristics and treatment, but also genetics and other omics, lifestyle (physical activity, nutrition, sleep, toxic habits), occupational exposures (including night shift work), environmental factors such as living area conditions, and medical history, aiming to build integrative prognosis models. This multidisciplinary study will provide a complete evaluation of the biological, clinical, environmental, lifestyle and socioeconomic factors determining survival of patients of the three cancers and of the inter-relationship of these factors. The following are the specific objectives for each cohort: For the colorectal cancer cohort: (1) to study the accomplishment of primary treatment with ESMO (European Society for Medical Oncology) and ASCO (American Society of Clinical Oncology) guidelines and factors associated with it; (2) to study factors associated with survivorship, response to treatment and toxicity due to chemotherapy using genetic, epidemiological and clinical-pathological variables; and (3) to validate those models via comparison with Glasgow Prognostic Score predictions. For the breast cancer cohort: (1) to study whether first-line treatment accomplished the St Gallen International Expert Consensus recommendations; (2) to study factors

associated with survivorship, response to treatment and toxicity due to chemotherapy using genetic, epidemiological and clinical-pathological variables; and (3) to validate those models via comparison with the Nottingham Prognostic Index and Adjuvant! For prostate cancer cohort: (1) to analyse the adequacy of initial treatment to the recommendations by the European Association of Urology and the National Institute for Health and Care Excellence; (2) to elaborate models on survivorship, risk of biochemical relapse, quality of life, response to primary treatment and toxicity to chemotherapy/brachytherapy; and (3) to validate survivorship and risk of biochemical relapse models via comparison with Han and Kattan nomograms. In this article, we report the study design, the main description of all three cohorts and the preliminary results on survival.

## COHORT DESCRIPTION AND METHODS

The MCC-Spain began as a case-control study in 2008, started by the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), on both genetic and environmental exposures associated with colorectal, female breast, prostate and gastric cancers and chronic lymphocytic leukaemia. Its design has been published elsewhere<sup>13</sup>; it recruited 10 183 incident cases and controls between 2008 and 2013 in 12 Spanish provinces (Asturias, Barcelona, Cantabria, Girona, Granada, Gipuzkoa, Huelva, León, Madrid, Murcia, Navarra and Valencia). Using the incident cases originally recruited between 2008 and 2013, and given that in 2016 the MCC-Spain has turned towards the identification of factors associated with cancer prognosis, inception cohorts on colorectal, breast and prostate cancers have been assembled, enrolling the patients for a prospective follow-up carried out in 2017–2018. From here on, we only refer to the recruited cases of colorectal (2140 cases), breast (1738 cases) and prostate (1112 cases) cancers; their distribution by province and hospital appears in online supplementary table 1 and the flow chart appears in figure 1.

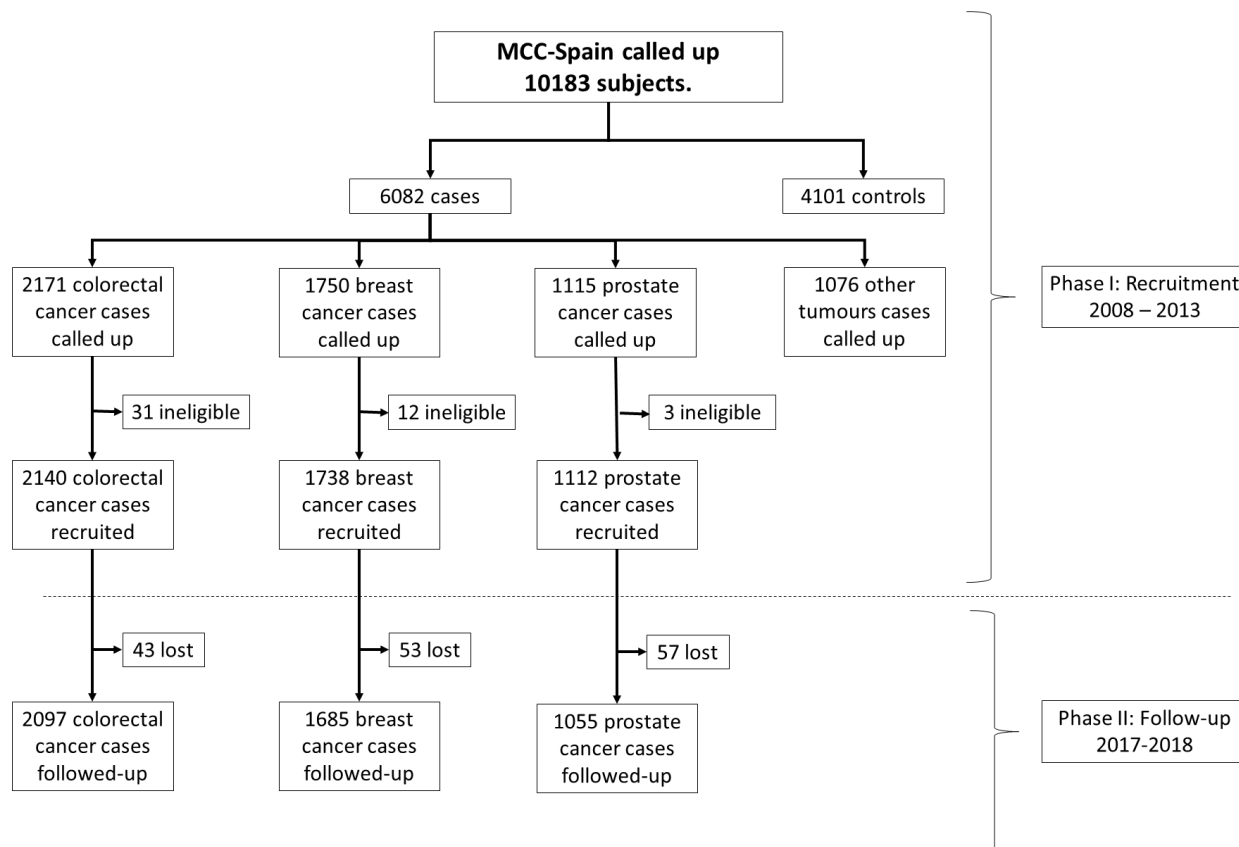
### Patient recruitment and public involvement statement

Patients recruited were between 20 and 85 years old, had resided in the catchment area for at least 6 months before the recruitment and were able to answer the epidemiological questionnaire, and had incident colorectal, breast or prostate cancer. For the recruitment, study personnel contacted newly diagnosed cancer cases in the 21 collaborating hospitals. Cases were identified as soon as possible after the diagnosis; only histologically confirmed incident cases were included.

Participants are being informed on the project's main results via flyers. There is no other patient involvement.

### Information at recruitment and biological samples

The information obtained and its timing is summarised in table 1.



**Figure 1** Flow chart of the participants in the MCC-Spain study. MCC-Spain, multicase-control study in Spain.

Information about sociodemographic, personal and familial medical history, use of drugs, reproductive history, physical activity, and environmental and occupational exposures was gathered using a standardised questionnaire<sup>14</sup> administered by trained personnel in a face-to-face interview. Diet information in the year before diagnosis was obtained using a validated semiquantitative frequency-food questionnaire<sup>15</sup> filled by the participants. Both questionnaires can be found at <http://www.mccspain.org>. Biological samples were obtained, including peripheral blood or saliva (from 92% of breast cancer cases, 95% of colorectal cancer cases and 97% of prostate cancer cases), toenail and hair (from 77% and 81% of participants, respectively), urine or tumour biopsies. Regarding peripheral blood, 27 mL was aliquoted in whole blood, plasma, serum and cellular fraction for DNA extraction and stored at  $-80^{\circ}\text{C}$ . Saliva was collected from people unable to donate a blood sample.

### Genotyping

From 80% of the participants, a genotype of exome was made using the Illumina Infinium HumanExome. In addition to the about 250 000 exome variants included in the original beadchip, 6000 SNPs previously found in GWAS (Genome-wide Association Study) or localised in metabolic pathways of interest were added on MCC-Spain researchers' request. MCC-Spain has recently obtained funding for carrying out a GWAS with all the participants

and to launch an analysis on circulant microRNA in patients with breast cancer.

### Initial clinical information

Trained personnel reviewed the medical records in order to collect information on pathology characteristics, tumour extension, clinical data, first-line treatment and recurrence. For colorectal cancer cases, we documented the first biopsy, tumour location, surgical piece dimensions, histological type according to the International Classification of Diseases for Oncology, Third Edition, tumour, node, metastases (TNM) status, carcinoembryonic antigen levels, and first-line treatment (surgery extension, if done; neoadjuvant, adjuvant or palliative chemotherapy or radiotherapy). For breast cancers, we obtained information on tumour location, differentiation's degree, immunohistochemical characteristics (hormonal receptors, Erb-B2), TNM status and first-line treatment (mastectomy/conservative surgery; neoadjuvant, adjuvant or palliative hormone-therapy, chemotherapy or radiotherapy; target-directed therapy such as trastuzumab). For prostate cancer cases, we gathered information on tumour location, Gleason score, D'Amico classification, TNM status, PSA levels and first-line treatment (none, surgery, hormone-therapy, chemotherapy or radiotherapy, including, when appropriate, the purpose of therapy—neoadjuvant, adjuvant or palliative). TNM status for all three tumours was classified according to TNM sixth edition.

**Table 1** Information obtained in the MCC-Spain

Phase	Measurements
Phase I: recruitment	<p>2008–2013</p> <p><i>Contact with newly diagnosed cancer cases.</i></p> <p><i>Trained personnel perform a structured computerised epidemiological questionnaire in a face-to-face interview to obtain the following information: sociodemographic, personal and familial medical history, use of drugs, reproductive history, physical activity, environmental and occupational exposures.</i></p> <p><i>A validated semiquantitative frequency-food questionnaire is self-completed to obtain diet information.</i></p> <p><i>Biological samples are obtained: peripheral blood or saliva, toenail, hair, urine, tumour biopsies.</i></p> <p><i>A genotype of exome is made using the Illumina Infinium HumanExome.</i></p> <p><i>Medical records review by trained personnel to obtain the following: pathology characteristics, tumour extension, clinical data, first-line treatment, recurrence.</i></p> <p>For colorectal cancer cases      First biopsy, surgical piece dimensions, histological type, carcinoembryonic antigen levels.</p> <p>For breast cancers cases      Differentiation's degree, immunohistochemical characteristics.</p> <p>For prostate cancer cases      Gleason score, D'Amico classification, PSA levels.</p>
Phase II: follow-up	<p>2017–2018</p> <p>Medical records review by trained personnel to obtain the following:</p> <p>For colorectal cancer cases      TNM status at recruitment, first-line treatment, surgical margins, patient status after first-line treatment, appearance of second primary tumour, current patient's vital status.</p> <p>For breast cancers cases      Histological grade at diagnosis, Nottingham index, complete clinical/pathological remission, grade of response to treatment, relapse, second primary tumour, current patient's vital status.</p> <p>For prostate cancer cases      PSA concentration, Gleason grade and biopsy characteristics at diagnosis, pathological characteristics of the surgical specimen, first-line treatment, clinical response to first-line treatment, second primary tumour, current patient's vital status.</p> <p>Consult in the IND to realise the vital status of patients.</p> <p>Contact by phone to complete specific quality of life questionnaires.</p> <p>For colorectal cancer cases      SF-12, FACT-Colorectal Symptom Index.</p> <p>For breast cancer cases      SF-12, FACT/NCCN Breast Symptom Index.</p> <p>For prostate cancer cases      SF-12, Charlson Comorbidity Index, FACT-P questionnaire, International Prostate Symptom Score.</p>

SF-12: 12-Item Short Form Survey; FACT: Functional Assessment of Cancer Therapy (P for prostate, B for breast cancer); FACT/NCCN: Functional Assessment of Cancer Therapy/National Comprehensive Cancer Network.

IND, Índice Nacional de Defunciones; MCC-Spain, multicase-control study in Spain; PSA, prostate-specific antigen; TNM, tumour, node, metastases.

### Follow-up information

Follow-up was carried out between 2017 and 2018 by reviewing medical records. For patients with colorectal cancer, we collected data on TNM status at recruitment, first-line treatment, surgical margins, patient status after first-line treatment (free of disease, partial response, progression, relapse or stable disease), appearance of second primary tumour and current patient's vital status. For patients with breast cancer, we gathered information on histological grade at diagnosis, Nottingham index,

complete clinical/pathological remission, grade of response to treatment (according to the Miller and Payne system or similar classifications), relapse, second primary tumour and current patient's vital status. For patients with prostate cancer, the information assembled included PSA concentration, Gleason grade and biopsy characteristics at diagnosis, pathological characteristics of the surgical specimen, first-line treatment, clinical response to first-line treatment (stable disease/progression or relapse/unknown), chemical relapses, relapse clinical

characteristics (local/metastatic and its location), second primary tumour, and current patient's vital status. Some of these data were obtained in order to double-check the clinical information collected at recruitment.

The National Death Index (Índice Nacional de Defunciones (IND)) was consulted to realise the vital status of patients whose last contact with the hospital had occurred 3 or more months before our revision of his/her medical record. The IND is a nationwide database supported by the Spanish Ministry of Health; it is intended to allow the researchers to establish the vital status of patients under study.<sup>16</sup>

Patients alive at follow-up were contacted by phone and asked to complete specific quality of life questionnaires: SF-12<sup>17</sup> (12-Items Short Form Survey; colorectal, breast and prostate cancers), FACT-Colorectal Symptom Index<sup>18</sup> (Functional Assessment of Cancer Therapy; colorectal cancer), FACT/NCCN (National Comprehensive Cancer Network) Breast Symptom Index<sup>19</sup> (breast cancer), and for prostate cancer the Charlson Comorbidity Index,<sup>20</sup> the FACT-P questionnaire (Functional Assessment of Cancer Therapy Prostate)<sup>21</sup> and the International Prostate Symptom Score.<sup>22</sup>

The number of patients with follow-up is 2097 for colorectal, 1685 for breast and 1055 for prostate cancer cohorts. This gives a 91% statistical power for colorectal cancer to detect an HR  $\geq 1.2$ ; an 83% statistical power for breast cancer to detect the same HR; and an 80% statistical power for prostate cancer to detect an HR  $\geq 1.25$  (assuming 20% exposed patients and 75%, 90% and 85% survival probability in the non-exposed group, respectively).

### Statistical analysis

For preliminary results shown in this paper, data are described using absolute frequencies with percentages and means with SD. Patients who died by any cause before the end of follow-up were classified as events and censored otherwise. Time of follow-up was the difference between date of diagnosis and date of death or date of last contact with the hospital or the researchers. Survival probabilities were obtained using unadjusted Kaplan-Meier estimators. Further analyses should deal with confounding and modifiers using multivariate regression models (eg, Cox or Weibull regression). Initial treatment could be related with both basal factors and survivorship, eventually leading to confounding by indication; it would be controlled using propensity scores.

### Ethics

The protocol of MCC-Spain was approved by the ethics committees of the participating institutions.<sup>13</sup> At recruitment, all participants were informed about the study objectives and signed an informed consent, which also included the authorisation for following up the patient via medical records or phone calls. Only participants agreeing to being followed up were included in the inception cohorts. Confidentiality of data is secured by

removing personal identifiers in the data sets. The database was registered in the Spanish Agency for Data Protection (number 2102672171).

### FINDINGS TO DATE

The MCC-Spain has provided results on the effects of different risk factors. For instance, night shift work increased the risk of more aggressive prostate cancers,<sup>23</sup> although this excess risk almost disappeared 20 years after last exposure<sup>24</sup>; long-term consumption of calcium channel blockers was associated with higher breast cancer risk in overweight women<sup>25</sup>; adherence to the Western dietary patterns increased breast cancer risk in both premenopausal and postmenopausal women<sup>26</sup>; first validation in a European population of a risk model for breast cancer developed in American women using both modifiable and non-modifiable risk factors as well as 92 genetic variants<sup>27</sup>; use of environmental and genetic factors to elaborate a model to stratify the risk of colorectal cancer<sup>28</sup>; and adherence to the World Cancer Research Fund/American Institute for Cancer Research nutrition-based guidelines was associated with lower risk of colorectal and breast cancers, but not of prostate cancer.<sup>29</sup> A complete list of published results from MCC-Spain appears in online supplementary table 2 and supplementary reference list.

Initial results of the follow-up are shown in this work. [Table 2](#) displays the main characteristics of the patients; [table 3](#) details specific information of each tumour; and [table 4](#) describes first-line treatment.

### Colorectal cancer

Out of 2140 patients with colorectal cancer, 2097 (98%) have been followed. They were  $67 \pm 10.9$  years old on average at recruitment, and 1334 (63.4%) were men. The first case was recruited on 18 March 2007 and the follow-up was closed on 23 August 2018, accounting for 12813.8 person-years of follow-up. During this period, 819 (39.1%) cases died. Linearised mortality rate was 6.4 per 100 patient-years (95% CI 6.0 to 6.8) ([table 2](#)).

Most cases (1882, 90%) were adenocarcinoma, and the most frequent location was rectum-sigma (37.7%) and the less frequent right colon (27%). Of the patients, 52% were at clinical stage II or lower, and in 110 patients (5.3%) we could not establish the clinical stage. Of the cancers, 52.5% were moderately differentiated (grade II) and 24.8% well differentiated (grade I) ([table 3](#)).

Surgery was carried out in 1999 patients with colorectal cancer, and it was for palliative purposes in 127 patients (6.1%). There were 169 patients who were treated via endoscopy, reaching complete resection in 107 of them. There were 1518 (72.4%) patients who received chemotherapy, and most of them (1451) were for adjuvant or neoadjuvant purposes; 488 (23.2%) received radiotherapy (401 neoadjuvant, 82 adjuvant and only 5 palliative) ([table 4](#)).

**Table 2** Main characteristics of the followed patients

Variable	Category	Colorectal cancer (n=2097)	Breast cancer (n=1685)	Prostate cancer (n=1055)
Age, mean ( $\pm$ SD)		66.98 ( $\pm$ 10.85)	56.5 ( $\pm$ 12.6)	65.86 ( $\pm$ 7.38)
Gender	Female	763 (36.39%)	1685 (100%)	–
	Male	1334 (63.61%)	–	1055 (100%)
Postmenopausal	Yes	–	1095 (65.0%)	–
	No	–	589 (35.0%)	–
	Missing	–	1 (0.1%)	–
Histology (specific types in each tumour)		Adenocarcinoma: 1882 (89.75%)	Ductal: 1276 (75.7%)	Adenocarcinoma (acinar): 1053 (99.91%)
		Mucinous adenocarcinoma: 125 (5.96%)	Lobular :110 (6.5%)	Others: 2 (0.09%)
		Signet ring cells adenocarcinoma: 12 (0.57%)	Paget disease: 19 (1.1%)	–
		Others: 4 (0.19%)	Others: 280 (16.6%)	–
		Unknown: 74 (3.53%)	–	–
Tumour size	T0	98 (4.67%)	23 (1.4%)	–
	T1	125 (5.96%)	861 (51.1%)	227 (21.52%)
	T2	283 (13.49%)	424 (25.2%)	521 (49.38%)
	T3	1172 (55.89%)	73 (4.3%)	98 (9.29%)
	T4	319 (15.21%)	39 (2.3%)	8 (0.76%)
	Tis	–	109 (6.5%)	–
	Missing	100 (4.77%)	156 (9.3%)	196 (18.58%)
	Not evaluable	–	–	5 (0.47%)
Node infiltration	N0	1193 (56.89%)	877 (52.0%)	271 (25.69%)
	N1	515 (24.56%)	441 (26.2%)	9 (0.85%)
	N2	286 (13.64%)	186 (11.0%)	–
	N3	–	5 (0.3%)	–
	Missing	103 (4.91%)	176 (10.4%)	224 (21.23%)
	Not evaluable	–	–	551 (52.23%)
Metastasis	No	1721 (82.07%)	1376 (81.7%)	532 (50.43%)
	Yes	330 (15.74%)	41 (2.4%)	17 (1.61%)
	Missing	46 (2.19%)	268 (15.9%)	215 (20.38%)
	Not evaluable	–	–	291 (27.58%)
Clinical stage	0	77 (3.67%)	–	–
	I	338 (16.12%)	702 (41.7%)	367 (34.79%)
	II	673 (32.09%)	479 (28.4%)	496 (47.01%)
	III	569 (27.13%)	179 (10.6%)	132 (12.51%)
	IV	330 (15.74%)	41 (2.4%)	17 (1.61%)
	Missing	110 (5.25%)	284 (16.9%)	43 (4.08%)

The 5-year survival probability estimated via Kaplan-Meier was 71.6% (95% CI 69.6 to 73.5) (figure 2A). Survival was higher in women (74.4%, 95% CI 71.0 to 77.2) than in men (70.0%, 95% CI 67.5 to 72.4) ( $p < 0.001$ ) (figure 2B). The 5-year survival probability was 85.2% (81.0–88.6) in patients diagnosed with stage I, 84.0%

(81.0–86.6) with stage II, 73.4% (69.6–76.9) with stage III and 27.6% (22.9–32.5) with stage IV (figure 3A).

### Breast cancer

The maximum span for breast cancer follow-up was 9.5 years (from 13 July 2007 to 22 March 2017). Follow-up

**Table 3** Specific information for each cancer

Specific information for colorectal cancer		Specific information for breast cancer		Specific information for prostate cancer	
Location	Right colon	566 (26.99%)	Progesterone receptor	1398 (83.0%)	1 (Gleason score=6)
	Left colon	719 (34.29%)	Her2	244 (14.5%)	2 (Gleason score=3+4)
	Rectum-sigma	791 (37.72%)	Intrinsic subtype	43 (2.6%)	3 (Gleason score=4+3)
	Unknown	21 (1%)		1237 (73.4%)	4 (Gleason score=8)
Differentiation's degree	I	520 (24.8%)	Estrogen receptor	401 (23.8%)	5 (Gleason score=9 or 10)
	II	1100 (52.46%)	Progesterone receptor	47 (2.8%)	Missing
	III	247 (11.78%)	Her2	294 (17.4%)	39 (3.70%)
	Not evaluable	230 (10.97%)	Basal-like	1250 (74.2%)	
			Luminal A	141 (8.4%)	
			Luminal B	997 (59.2%)	
		Her2	331 (19.6%)		
		Basal-like	81 (4.8%)		
		Luminal ONI	130 (7.7%)		
		Non-luminal ONI	91 (5.4%)		
		Missing	13 (0.8%)		
		I	42 (2.5%)		
		II	329 (19.5%)		
		III	520 (30.9%)		
		Missing	355 (21.1%)		
		Grade	481 (28.5%)		
		PSA (ng/mL)			
		D'Amico			
		Low risk			
		Intermediate risk			
		High risk			
		Missing			

ONI, otherwise not identified; PSA, prostate-specific antigen.

**Table 4** First-line treatment

Treatment	Category	Colorectal cancer	Breast cancer	Prostate cancer
None (active surveillance)		–	–	38 (3.6%)
Surgery		Total: 1999 (95.3%) Resection: 1800 (85.8%) Palliative: 127 (6.1%) No resection: 61 (2.9%) Others: 11 (0.5%)	Conservative: 1231 (73.1%) Mastectomy: 454 (26.9%)	Prostatectomy: 639 (61.4%)
Chemotherapy	Neoadjuvant	427 (20.4%)	200 (11.9%)	1 (0.1%)
	Adjuvant	1024 (48.8%)	664 (39.4%)	1 (0.1%)
	Palliative	67 (3.2%)	25 (1.5%)	7 (0.7%)
Radiotherapy	Neoadjuvant	401 (19.1%)	5 (0.3%)	227 (21.5%)
	Adjuvant	82 (3.9%)	1132 (67.2%)	36 (3.4%)
	Palliative	5 (0.2%)	21 (1.2%)	2 (0.2%)
Endocrine therapy	Yes	–	1023 (60.7%)	Adjuvant to surgery: 19 (1.8%) Adjuvant to radiotherapy: 99 (9.4%) Neoadjuvant: 102 (9.7%) Palliative: 69 (6.5%)
	No	–	662 (39.3%)	689 (65.3%)
Others (specify for each tumour)	Endoscopy	Complete resection: 107 (5.1%) Non-complete resection: 62 (3.0%)	–	–
	Her2-targeted therapy	–	152 (9.0%)	–
	Cryotherapy	–	–	21 (2.0%)
	Transurethral resection	–	–	4 (0.4%)

was obtained for 1685 out of 1738 patients with breast cancer (97%), adding 10931 person-years; 206 patients died in the follow-up. The linearised mortality rate was 1.9 per 100 patient-years (95% CI 1.6 to 2.2).

Women with breast cancer were 56.5±12.6 years old on average at recruitment, and 65% were postmenopausal. The most usual type of tumour was ductal (75.7%), followed by lobular (6.5%). Most breast cancers were diagnosed at early stages (71% at stages I or II) and only 41 (2.4%) had metastasised at the time of diagnosis (table 2). Of the cancers, 83% were oestrogen receptor-positive, 73.4% progesterone receptor-positive and 17.4% Her2-positive. Regarding intrinsic subtypes, 997 (59.2%) could be classified as luminal A, 331 (19.6%) as luminal B, 81 (4.8%) as Her2 and 130 (7.7%) as basal-like. According to grade of differentiation, moderately differentiated accounted for 30.9% of breast cancers, and well differentiated and bad differentiated accounted for about 20% of cancers each. Grade could not be obtained from medical records in 481 patients (28.5%) (table 3).

Conservative surgery was performed in 1231 (73.1%) patients and mastectomy in the remaining 454 (26.9%). Adjuvant or neoadjuvant chemotherapy was administered to 50.3% of patients, while radiotherapy was used in 1158 women (68.7%), endocrine therapy was used in 1023 women (60.7%) and Her2-targeted therapy in 152 patients (9.0%) (table 4). Kaplan-Meier 5-year survival

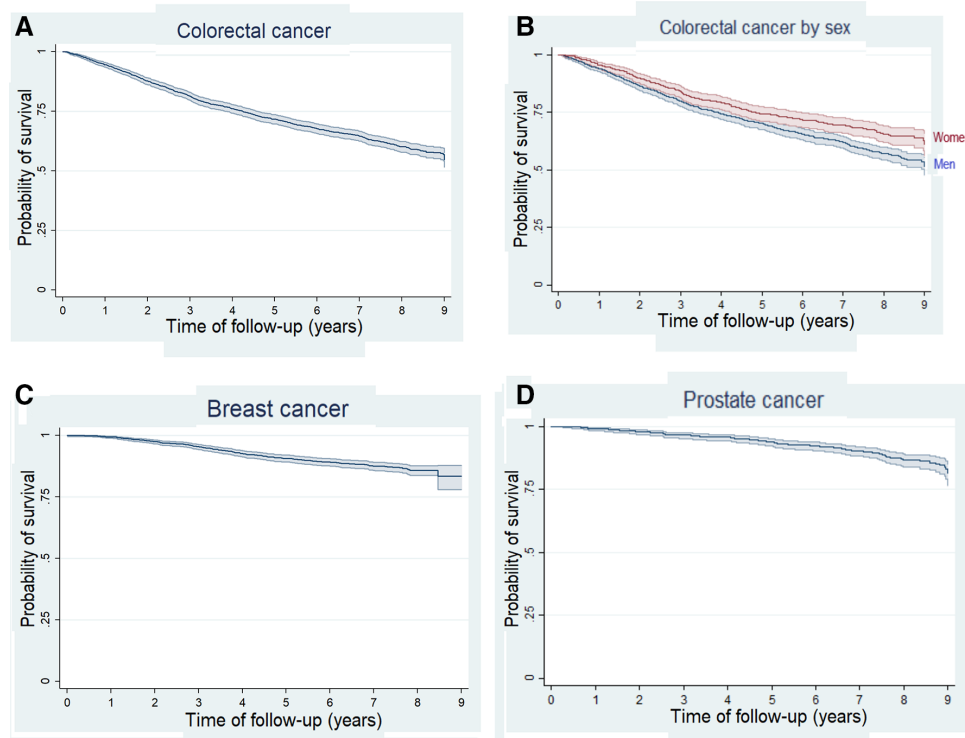
with breast cancer was 90.7% (95% CI 89.2 to 92.0) (figure 2C). Women diagnosed with stage I had 97% (95.5–98.1) 5-year survival probability, 91.9% (89.1–94.1) at stage II, 84.1% (77.8–88.7) at stage III and 38.5% (18.6–58.2) at stage IV (figure 3B).

### Prostate cancer

A total of 1112 men with prostate cancer were recruited and 1055 (94.9%) have been followed up. The first patient was included on 26 January 2008 and the end of follow-up was on 13 July 2018, adding 7169.6 person-years of follow-up. Patients were 65.9 years old on average at recruitment. There were 119 patients who died in the follow-up, making the linearised mortality rate 1.7 per 100 patient-years (95% CI 1.4 to 2.0).

Almost all prostate cancers (99.9%) were adenocarcinoma; 496 (47%) were diagnosed at stage II and 132 (12.5%) at stage III (table 2). The level of PSA gives an average of 11.5±16.3 ng/mL. Considering the Gleason score, 42.6% of prostate cancers were well differentiated (Gleason grade=1, ie, Gleason score=6), 28.3% were at Gleason grade 2 (Gleason score=3+4), and only 14.0% were bad differentiated (Gleason grade 4 or 5, ie, Gleason score ≥8). Gleason grade could not be established in 17.4% of patients. D'Amico classification system results in 31.4% of patients with low-risk cancer, 41.1% intermediate-risk and 27.4% high-risk cancer (table 3).





**Figure 2** Kaplan-Meier survival estimates for colorectal cancer (A), colorectal cancer by sex (B), breast cancer (C) and prostate cancer (D).

Thirty-eight patients with prostate cancer were not treated medically at the beginning, being followed by active surveillance. Prostatectomy was performed in 61.4% cases, radiotherapy in 265 patients (25.1%) and endocrine therapy in 289 patients (27.4%). A small number of patients were treated via transurethral resection, cryotherapy or chemotherapy (table 4). The 5-year survival probability by Kaplan-Meier was 93.7% (95% CI 92.0 to 95.1) (figure 2D). Survival probability 5 years after being diagnosed was 94.5% (88.1–97.5) for patients at stage I, 95.6% (93.3–97.2) at stage II, 92.4% (88.5–95.0) at stage III and 70.5% (42.8–88.6) at stage IV (figure 3C).

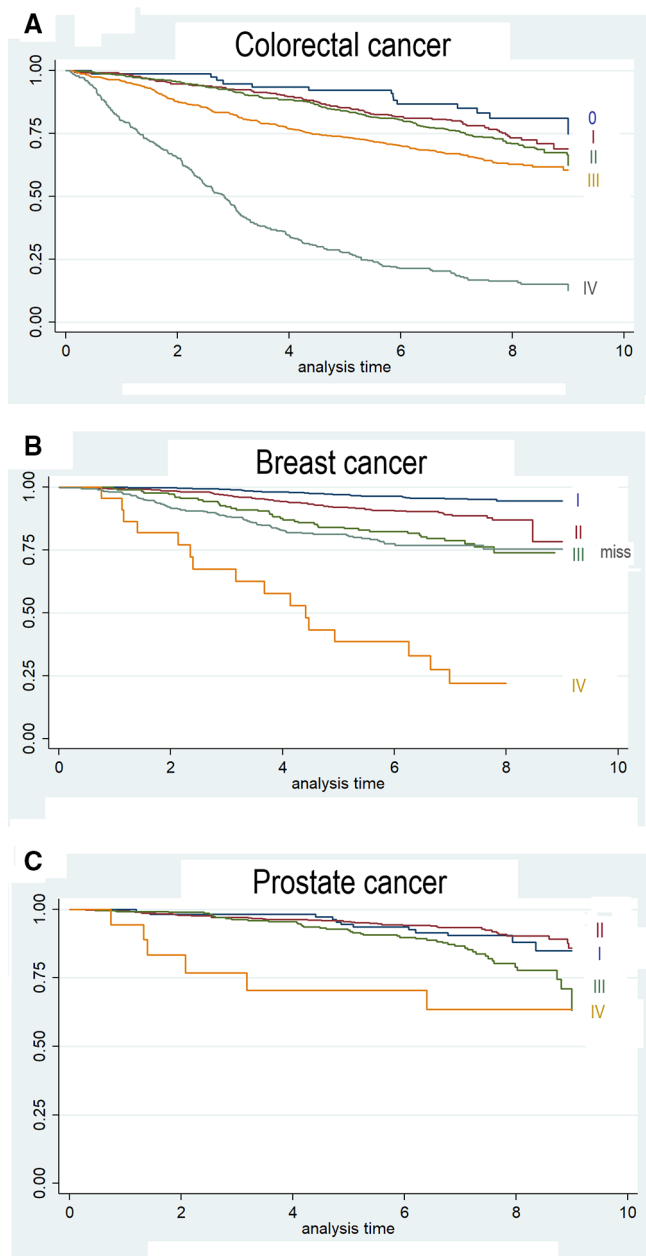
### STRENGTHS AND LIMITATIONS

In this article, we have described how three prospective cohorts on colorectal, breast and prostate cancers have been assembled from patients originally recruited for a case-control study, with 97% patients followed up and accounting for more than 30 000 person-years. This is a main achievement of a network settled within the CIBERESP in 12 Spanish provinces. The study is population-based and included only incident cancers. The amount of detailed information recorded as well as the availability of biological samples at recruitment will allow the identification of genetics, environmental, lifestyle and clinical prognosis factors in three frequent cancers in Spain. In this regard, a remarkable feature of the study is the feasibility of studying cancer risk factors as putative prognosis factors. For example, risk factors already analysed in the case-control phase were diet, circadian cycle

disruption, some drugs, endocrine disruptors, artificial light or proximity to green spaces; information regarding these risk factors was recorded at recruitment and is available for a prognosis factor analysis in the follow-up (see online supplementary material for a complete reference list of MCC-Spain articles).

Obtaining information on personal history, occupational exposures, diet, physical exercise or other lifestyle components is somewhat subjective as both patients and interviewers could be prone to be influenced by their feelings or beliefs about the hypotheses under study, eventually leading to differential misclassification bias. This could hardly have occurred in this study. First, patients were not aware of the hypotheses. Second, interviewers were familiar with the case-control study, not with the cohort design as it was decided later; therefore, if interviewers or patients have introduced some misclassification, it could probably have been non-differential, eventually leading to bias towards the null,<sup>30</sup> which would make more robust the positive findings in this cohort study.

This study also has some weaknesses. First, multi-centre studies are double-edged; they are needed in order to include many patients, but they could introduce heterogeneity in both the information gathered and the way patients are treated. In this regard, the analysis of prognosis factors should be adjusted for the hospital of recruitment. Second, 113 participating patients have been lost (43 with colorectal cancer, 53 with breast cancer and 57 with prostate cancer). We have tried to minimise it by searching information in three ways: medical



**Figure 3** Kaplan-Meier estimates by stage at diagnosis for colorectal cancer (A), breast cancer (B) and prostate cancer (C).

records, phone calls and IND; however, we cannot rule out that some patients without follow-up could have died. It is noteworthy that, due to the small number of patients without follow-up, the maximum bias it could introduce in our survival estimates is 2% for colorectal cancer, 3% for breast cancer and 5% for prostate cancer. Third, we have not obtained information on lifestyle changes after diagnosis, which limits lifestyle analysis to habits before cancer appearance. Fourth, the number of patients included in our cohorts is small compared with those based on cancer registries, limiting the analysis of subgroups.

Summarising, the MCC-Spain study has assembled three cohorts with about 4700 patients with cancer accounting for 30 000 patient-years of follow-up, with only 3% patient

withdrawals. The information gathered at recruitment will allow to prospectively investigate clinical, lifestyle, environmental and genetic variables as prognosis factors in colorectal, breast and prostate cancers in Spain.

### Samples

Biological samples were stored at the biobanks supported by Instituto de Salud Carlos III-FEDER: Parc de Salut MAR Biobank (MARBiobanc) (RD09/0076/00036), 'Biobanco La Fe' (RD 09 0076/00021) and FISABIO Biobank (RD09 0076/00058), and also at the Public Health Laboratory from Gipuzkoa, the Basque Biobank, the ICOBIOBANC (sponsored by the Catalan Institute of Oncology), the IUOPA Biobank from the University of Oviedo and the ISCIII Biobank.

### Collaborators

MCC-Spain already participates in international consortiums such as Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO; <https://www.fredhutch.org/en/labs/phs/projects/cancer-prevention/projects/gecco.html>), Breast Cancer Association Consortium (BCAC; <http://bcac.ccge.medschl.cam.ac.uk/>) and Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL; <http://practical.icr.ac.uk/blog/>), where MCC-Spain would contribute to study interactions among the putative prognosis factors in vast population samples.

### Author affiliations

- <sup>1</sup>University of Cantabria – IDIVAL, Santander, Spain
- <sup>2</sup>Grupo de Investigación en Interacciones Gen-Ambiente y Salud (GIIGAS), Instituto de Biomedicina (IBIOMED), Universidad de León, León, Spain
- <sup>3</sup>CIBER Epidemiología y Salud Pública (CIBERESP), Granada, Spain
- <sup>4</sup>Instituto de Investigación Biosanitaria de Granada (ibs.GRANADA), Granada, Spain
- <sup>5</sup>Cancer and Environmental Epidemiology Unit, National Center for Epidemiology, Carlos III Institute of Health, Madrid, Spain
- <sup>6</sup>Cancer Epidemiology Research Group, Oncology and Hematology Area, IIS Puerta de Hierro (IDIPHIM), Madrid, Spain
- <sup>7</sup>Cancer Prevention and Control Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Spain
- <sup>8</sup>Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain
- <sup>9</sup>Public Health Division of Gipuzkoa, BioDonostia Research Institute, San Sebastian, Spain
- <sup>10</sup>Navarra Public Health Institute, Pamplona, Spain
- <sup>11</sup>IdiSNA, Navarra Institute for Health Research, Pamplona, Spain
- <sup>12</sup>Oncology Institute, University of Oviedo, Oviedo, Spain
- <sup>13</sup>Centro de Investigación en Recursos Naturales, Salud y Medio Ambiente (RENSMA), Universidad de Huelva, Huelva, Spain
- <sup>14</sup>Área de Cáncer y Salud Pública, FISABIO-Salud Pública, Valencia, Spain
- <sup>15</sup>Unitat d'Epidemiologia i Registre de Càncer de Girona (UERCG), Pla Director d'Oncologia, Institut Català d'Oncologia, Institut d'Investigació Biomèdica de Girona (IdIBGi), Universitat de Girona, Girona, Spain
- <sup>16</sup>Department of Epidemiology, Regional Health Authority, IMIB-Arrixaca, Murcia University, Murcia, Spain
- <sup>17</sup>Epidemiology Section, Public Health Division, Department of Health, Madrid, Spain
- <sup>18</sup>ISGlobal, Barcelona, Spain
- <sup>19</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain
- <sup>20</sup>IMIM (Hospital Del Mar Medical Research Institute), Barcelona, Spain

**Acknowledgements** We thank all the subjects who participated in the study and all MCC-Spain collaborators. SNP genotyping services were provided by the Spanish 'Centro Nacional de Genotipado' (CEGEN-ISCIII).

**Contributors** JA, AJM, JJJ-M have contributed to the conception and design of the study, considering the same contribution. JAM, AJM, JJJ-M, BP-G, VMA, VMO, PA, EA, SdS, IS, GF-T, JA, DS, RM-G, MDC, NA, GC-V, MP, MK and JL have acquired the data and have been involved in drafting the manuscript. All of them read and approved the final manuscript.

**Funding** The study was partially funded by the 'Accion Transversal del Cancer', approved on the Spanish Ministry Council on 11 October 2007, by the Instituto de Salud Carlos III-FEDER (PI08/1770, PI08/0533, PI08/1359, PS09/00773-Cantabria, PS09/01286-León, PS09/01903-Valencia, PS09/02078-Huelva, PS09/01662-Granada, PI11/01403, PI11/01889-FEDER, PI11/00226, PI11/01810, PI11/02213, PI12/00488, PI12/00265, PI12/01270, PI12/00715, PI12/00150, PI14/01219, PI14/0613, PI15/00069, PI15/00914, PI15/01032, PI17CIII/00034, PI18/00181), by the Fundación Marqués de Valdecilla (API 10/09), by the ICGC International Cancer Genome Consortium CLL (the ICGC CLL-Genome Project is funded by Spanish Ministerio de Economía y Competitividad (MINECO) through the Instituto de Salud Carlos III (ISCIII) and Red Temática de Investigación del Cáncer (RTICC) del ISCIII (RD12/0036/0036)), by the Junta de Castilla y León (LE22A10-2), by the Consejería de Salud de la Junta de Andalucía (PI-0571-2009, PI-0306-2011, salud201200057018tra), by the Conselleria de Sanitat of the Generalitat Valenciana (AP\_061/10), by the Recercaixa (2010ACUP 00310), by the Regional Government of the Basque Country, by the Consejería de Sanidad de la Región de Murcia, by the European Commission (grants FOOD-CT-2006-036224-HIWATE), by the Spanish Association Against Cancer (AECC) Scientific Foundation (GCTRA18022MORE), by the Catalan Government-Agency for Management of University and Research Grants (AGAUR) (grants 2017SGR723 and 2014SGR850), by the Fundación Caja de Ahorros de Asturias, and by the University of Oviedo. ISGlobal is a member of the CERCA Programme, Generalitat de Catalunya.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Jessica Alonso-Molero <http://orcid.org/0000-0002-1939-8798>

Beatriz Pérez-Gómez <http://orcid.org/0000-0002-4299-8214>

## REFERENCES

- Saadatmand S, Bretveld R, Siesling S, *et al*. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173 797 patients. *BMJ* 2015;351.
- Shukla N, Hagenbuchner M, Win KT, *et al*. Breast cancer data analysis for survivability studies and prediction. *Comput Methods Programs Biomed* 2018;155:199–208.
- Mirza AN, Mirza NQ, Vlastos G, *et al*. Prognostic factors in node-negative breast cancer. *Ann Surg* 2002;235:10–26.
- Zhang Z-yu, Luo Q-feng, Yin X-wei, *et al*. Nomograms to predict survival after colorectal cancer resection without preoperative therapy. *BMC Cancer* 2016;16:1–21.
- Merriell SWD, May MT, Martin RM. Predicting prostate cancer progression: protocol for a retrospective cohort study to identify prognostic factors for prostate cancer outcomes using routine primary care data. *BMJ Open* 2018;8:1–5.
- Riboli E, Hunt KJ, Slimani N, *et al*. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5:1113–24.
- Legendijk M, van Maaren MC, Saadatmand S, *et al*. Breast conserving therapy and mastectomy revisited: breast cancer-specific survival and the influence of prognostic factors in 129,692 patients. *Int. J. Cancer* 2018;142:165–75.
- Cardwell CR, Hicks BM, Hughes C, *et al*. Statin use after colorectal cancer diagnosis and survival: a population-based cohort study. *J Clin Oncol* 2014;32:3177–83.
- Petersson A, Robinson D, Garmo H, *et al*. Age at diagnosis and prostate cancer treatment and prognosis: a population-based cohort study. *Ann Oncol* 2018;29:377–85.
- Leone JP, Leone J, Zwenger AO, *et al*. Prognostic significance of tumor subtypes in women with breast cancer according to stage. *Am J Clin Oncol* 2019;42:588–95.
- Li Y, Feng Y, Dai W, *et al*. Prognostic effect of tumor sidedness in colorectal cancer: a SEER-Based analysis. *Clin Colorectal Cancer* 2019;18:e104–16.
- Roy S, Morgan SC. Who dies from prostate cancer? an analysis of the surveillance, epidemiology and end results database. *Clin Oncol* 2019;31:630–6.
- Castañó-Vinyals G, Aragonés N, Pérez-Gómez B, *et al*. Population-Based multicase-control study in common tumors in Spain (MCC-Spain): rationale and study design. *Gaceta Sanitaria* 2015;29:308–15.
- Estudio MCC-Spain. Epidemiological questionnaire, 2010. Available: [http://www.mccspain.org/wp-content/uploads/2016/07/Quest\\_MCCSpain.pdf](http://www.mccspain.org/wp-content/uploads/2016/07/Quest_MCCSpain.pdf) [Accessed 6 Sep 2019].
- Estudio MCC-Spain. Semi-Quantitative frequency-food questionnaire, 2010. Available: [http://www.mccspain.org/wp-content/uploads/2016/04/03\\_Cuestionario-alimentario\\_09Nov09.pdf](http://www.mccspain.org/wp-content/uploads/2016/04/03_Cuestionario-alimentario_09Nov09.pdf) [Accessed 6 Sep 2019].
- Navarro C. El Índice Nacional de Defunciones: un avance en la accesibilidad de los datos de mortalidad largamente esperado. *Gaceta Sanitaria* 2006;20:421–3.
- Ware J, Kosinski M, Turner-Bowker D, *et al*. How to score version 2 of the SF-12 health survey. *Lincoln, RJ Qual Inc* 2004.
- Colwell HH, Mathias SD, Turner MP, *et al*. Psychometric evaluation of the fact colorectal cancer symptom index (FCSI-9): reliability, validity, responsiveness, and clinical meaningfulness. *Oncologist* 2010;15:308–16.
- García SF, Rosenbloom SK, Beaumont JL, *et al*. Priority symptoms in advanced breast cancer: development and initial validation of the National comprehensive cancer Network-Functional assessment of cancer Therapy-Breast cancer symptom index (NFBISI-16). *Value Heal* 2012;15:183–90.
- Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
- Esper P, Mo F, Chodak G, *et al*. Measuring quality of life in men with prostate cancer using the funcionale assessment of cancer therapy-prostate instrument. *Adult Urol* 1997;30:920–8.
- Barry MJ, Fowler FJ, O'Leary MP, *et al*. The American urological association symptom index for benign prostatic hyperplasia. The measurement Committee of the American urological association. *J Urol* 1992;148:1549–57.
- Papantoniou K, Castañó-Vinyals G, Espinosa A, *et al*. Night shift work, chronotype and prostate cancer risk in the MCC-Spain case-control study. *Int J Cancer* 2015;137:1147–57.
- Kogevinas M, Espinosa A, Papantoniou K, *et al*. Prostate cancer risk decreases following cessation of night shift work. *Int J Cancer* 2019;145:2597–9.
- Gómez-Acebo I, Dierssen-Sotos T, Palazuelos C, *et al*. The use of antihypertensive medication and the risk of breast cancer in a case-control study in a Spanish population: the MCC-Spain study. *PLoS One* 2016;11:e0159672–14.
- Castelló A, Boldo E, Pérez-Gómez B, *et al*. Adherence to the Western, Prudent and Mediterranean dietary patterns and breast cancer risk: MCC-Spain study. *Maturitas* 2017;103:8–15.
- Dierssen-Sotos T, Gómez-Acebo I, Palazuelos C, *et al*. Validating a breast cancer score in Spanish women. The MCC-Spain study. *Sci Rep* 2018;8:1–8.
- Ibáñez-Sanz G, Díez-Villanueva A, Alonso MH, *et al*. Risk model for colorectal cancer in Spanish population using environmental and genetic factors: results from the MCC-Spain study. *Sci Rep* 2017;7:43263.
- Romaguera D, Gracia-Lavedan E, Molinuevo A, *et al*. Adherence to nutrition-based cancer prevention guidelines and breast, prostate and colorectal cancer risk in the MCC-Spain case-control study. *Int. J. Cancer* 2017;141:83–93.
- Hill HA, Kleinbaum DG. Encyclopedia of epidemiologic methods. In: Gail MH, Benichou J, eds. *Encyclopedia of epidemiologic methods*. Wiley, 2000: 92–3.